

Learning Physically Consistent Spectrogram-Based Representations for Underwater Acoustic Communications with Privileged Information

Luis Paulo Albuquerque Guedes

Abstract—This work proposes a physics-guided representation-learning framework for underwater acoustic communications using privileged information available only during training. A descriptor-space teacher encodes physical channel descriptors from paired transmitted-signal (TX) and received-signal (RX) acquisition cycles, and its latent representation is distilled into a convolutional neural network operating only on RX spectrograms. The student therefore supports blind RX-only inference without the transmitted waveform. Experiments on the SFI Smart Ocean Dataset for Acoustic Communications across receivers and frequency bands show that distillation preserves channel-related structure and improves transfer under receiver and waveform-identity shifts. Seen and unseen waveform identities are mapped to a shared coarse signaling-family space, avoiding comparisons between disjoint raw waveform classes.

Keywords—Underwater acoustic communications, deep representation learning, learning using privileged information, knowledge distillation, physics-guided learning.

I. INTRODUCTION

Underwater acoustic (UWA) communications support maritime monitoring, environmental sensing, autonomous platforms, and subsea operations. Their performance is constrained by multipath, reverberation, Doppler effects, and limited bandwidth, which produce time-varying and frequency-selective channels [1], [2]. Propagation models and simulators such as BELLHOP may not fully reproduce this operational variability [3], [4].

The SFI Smart Ocean Dataset for Acoustic Communications (SODAC) provides *in situ* received-signal (RX) recordings and synchronized transmitted-signal (TX) references for part of the waveform set [5]. This enables TX–RX information to guide training while retaining RX-only deployment.

We propose a physics-guided framework combining Learning Using Privileged Information (LUPI) and knowledge distillation. A teacher encodes physical descriptors extracted from paired TX–RX cycles, while a convolutional student learns physically informed representations directly from RX spectrograms. After training, the student operates without TX information.

The main contributions are: (i) a LUPI framework using paired TX–RX descriptors; (ii) a distilled RX-only encoder; (iii) cycle-disjoint evaluation across receivers, bands, and unseen waveform identities; and (iv) separate paired-signal diagnostics and non-cooperative RX-only evaluation.

Section II reviews related work; Section III describes SODAC and the input representation; Sections IV and V present the

method and evaluation protocol; Section VI discusses the results; and Section VII concludes the paper.

II. BACKGROUND AND RELATED WORK

UWA channels are affected by multipath, Doppler, and temporal dispersion [1], [2]. Classical approaches use propagation models [4] or TX-dependent descriptors [2], [3], limiting deployment without TX–RX pairing or environmental information.

Spectrogram-based deep learning has been applied to acoustic and UWA analysis [6], [7], [8], [13], but purely data-driven representations may transfer poorly and offer limited physical interpretability.

The proposed method combines LUPI [9] and knowledge distillation [10], [11]. Physical descriptors from paired *in situ* observations supervise an RX-only representation deployable without TX information. Related work includes modular underwater acoustic pipelines and side-information-based advantage distillation [14], [15].

III. DATASET AND INPUT REPRESENTATION

A. SODAC acquisition setup

SODAC was collected in a fjord in southwestern Norway [5]. The setup includes one transmitter at approximately 19–20 m depth and two receivers: Receiver 1 (R1), at 43 m depth and 0.27 km from the transmitter, and Receiver 2 (R2), at 14 m depth and 1.02 km away.

The dataset contains low-frequency (LF), medium-frequency (MF), and high-frequency (HF) bands, each with 80 acquisition cycles sampled at 96 kHz and transmissions every 10 minutes. Waveforms include LFM, PRBS, OFDM, OSDM, MFSK, JANUS, BCSK, FH-BCSK, and S2C-QPSK.

Synchronized TX references are available for PRBS, OFDM, OSDM, and MFSK. JANUS, BCSK, FH-BCSK, and S2C-QPSK are treated as unpaired RX recordings, as described in Section V.

Fig. 1 shows representative spectrograms and their time–frequency structures.

B. Spectro-temporal input representation

The received signal is modeled as

$$r(t) = (s * h)(t) + n(t), \quad (1)$$

where $s(t)$ is the transmitted signal, $h(t)$ is the channel impulse response, $*$ denotes convolution, and $n(t)$ is additive noise. The short-time Fourier transform (STFT) is

$$X(f, \tau) = \sum_t r(t) w(t - \tau) e^{-j2\pi f t}, \quad (2)$$

where $w(\cdot)$ is a Hann window.

All experiments use $N_{\text{FFT}} = 1024$, $N_w = 1024$, and hop size $H = 256$. The magnitude $|X(f, \tau)|$ is converted to dB, cropped to the corresponding band, and resized to (F, T) :

$$\mathbf{X} \in \mathbb{R}^{1 \times F \times T}. \quad (3)$$

Frame-wise normalization is avoided to preserve channel-related amplitude information. Spectrograms are standardized using training-partition statistics and remain compatible with two-dimensional convolutional neural networks [6], [16].

IV. PROPOSED METHOD

The method comprises three stages: (i) teacher learning from privileged physical descriptors; (ii) distillation into an RX-only student encoder; and (iii) blind RX-only inference under receiver and waveform shifts. Fig. 2 summarizes the training and inference pipeline.

A. Notation and physical descriptor vector

Let $p_j \in \mathbb{R}$ denote the j th scalar descriptor and

$$\mathbf{p} = [p_1, p_2, \dots, p_D]^\top \in \mathbb{R}^D \quad (4)$$

the complete descriptor vector. Throughout the paper, *descriptor* denotes a scalar quantity, whereas *descriptor vector* denotes $\mathbf{p}$.

The vector contains five quantities extracted from synchronized TX–RX cycles:

$$\mathbf{p} = [\text{peak}, \text{cfo}, \text{ds_rms}, \text{tail_50ms}, \text{env_kurtosis}]^\top. \quad (5)$$

The descriptors represent TX–RX coherence, carrier-frequency offset, RMS delay spread, normalized residual energy after 50 ms, and RX-envelope kurtosis, respectively. All descriptors are standardized using training-set statistics.

Because they require a known TX reference and synchronization, these descriptors are used as privileged information during training but are not required by the deployed RX-only student.

B. Descriptor-space teacher

The teacher is a multilayer perceptron (MLP) with two 128-unit hidden layers and rectified linear unit (ReLU) activations. Its encoder maps the standardized descriptor vector to

$$\mathbf{z}_T = g_T(\mathbf{p}), \quad \mathbf{z}_T \in \mathbb{R}^K, \quad (6)$$

and the reconstruction head produces

$$\hat{\mathbf{p}} = r_T(\mathbf{z}_T). \quad (7)$$

Training minimizes

$$\mathcal{L}_T = \|\mathbf{p} - \hat{\mathbf{p}}\|_1. \quad (8)$$

The ℓ_1 loss limits the influence of large residuals. After training, the reconstruction head is discarded and $\mathbf{z}_T$ becomes the privileged target.

C. RX-only student and knowledge distillation

The student is a compact two-dimensional CNN that maps an RX spectrogram to

$$\mathbf{z}_S = f_S(\mathbf{X}), \quad \mathbf{z}_S \in \mathbb{R}^K. \quad (9)$$

It contains four convolutional blocks with 32, 64, 128, and 128 channels, each comprising a 3×3 convolution, batch normalization, ReLU, and 2×2 max pooling. Global average pooling and a linear layer produce the embedding.

Teacher and student use $K = 128$, selected by validation-only grid search. This dimension provides a shared latent space and achieved the best validation performance within the evaluated range.

The student minimizes cosine distance to the frozen teacher target:

$$\mathcal{L}_{\text{dist}} = 1 - \frac{\mathbf{z}_S^\top \mathbf{z}_T}{\|\mathbf{z}_S\|_2 \|\mathbf{z}_T\|_2 + \varepsilon}, \quad (10)$$

where $\varepsilon > 0$ ensures numerical stability. Cosine alignment transfers latent direction while remaining insensitive to magnitude differences between modalities.

The mean-squared-error alternative is

$$\mathcal{L}_{\text{MSE}} = \frac{1}{K} \|\mathbf{z}_S - \mathbf{z}_T\|_2^2, \quad (11)$$

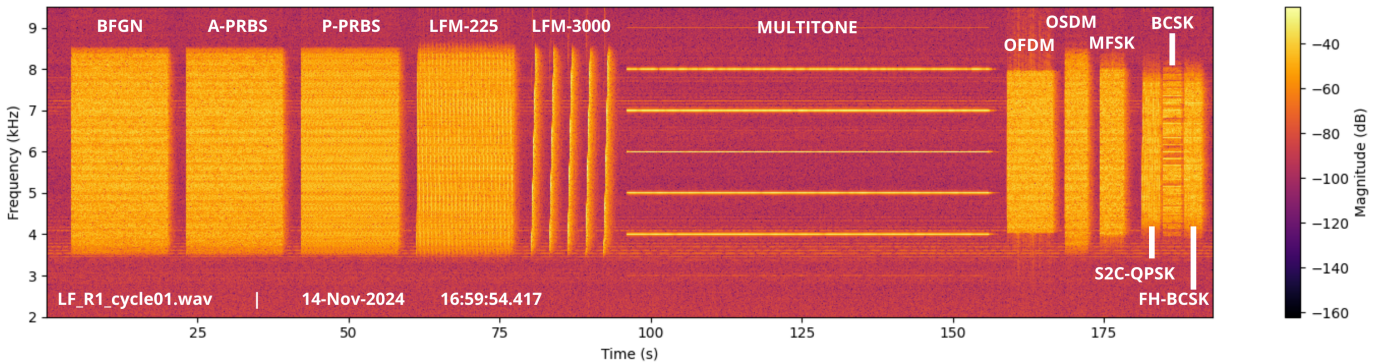


Fig. 1. Representative short-time Fourier transform spectrograms from the SODAC LF band at R1. Magnitude is expressed in dB without frame-wise normalization.

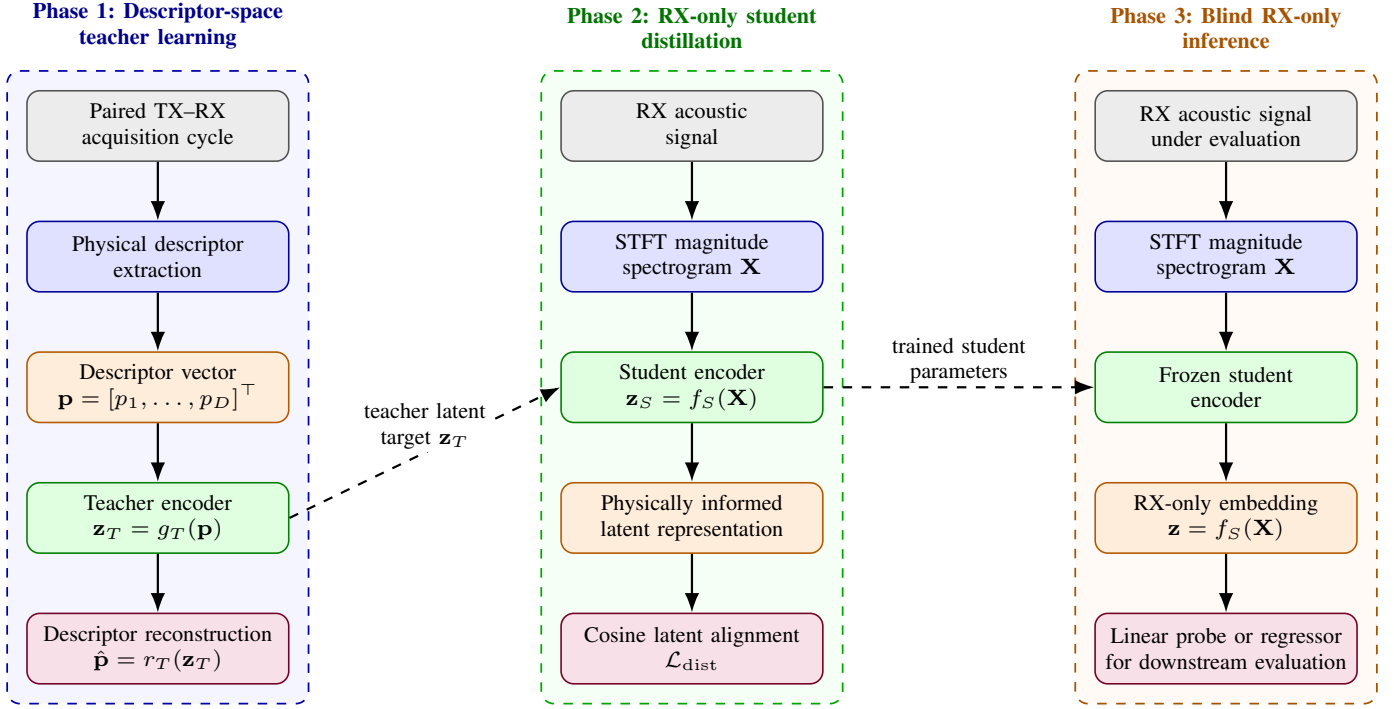


Fig. 2. Proposed privileged-information framework. Phase 1 trains a teacher on physical descriptors from paired TX-RX cycles. Phase 2 distills the teacher latent representation into an RX-only convolutional student through cosine alignment. Phase 3 discards the teacher and TX-side information and evaluates the frozen student on RX spectrograms under receiver and unseen-waveform shifts.

and the hybrid objective is

$$\mathcal{L}_{\text{hybrid}} = \lambda_{\text{hyb}} \mathcal{L}_{\text{dist}} + (1 - \lambda_{\text{hyb}}) \mathcal{L}_{\text{MSE}}, \quad (12)$$

with $\lambda_{\text{hyb}} = 0.5$.

with $\lambda * \text{hyb} = 0.5$.

D. Blind RX-only inference

After distillation, the teacher, reconstruction head, descriptors, and TX-side information are discarded. The frozen student produces

$$\mathbf{z} = f_S(\mathbf{X}) \quad (13)$$

from RX spectrograms alone. Linear regressors evaluate descriptor information, and linear probes evaluate coarse signaling-family recognition using frozen embeddings.

V. EVALUATION PROTOCOL

A. Cycle-disjoint partitions

To prevent temporal leakage, acquisition cycles are split by band and receiver into 60% training, 20% validation, and 20% test subsets, with waveform stratification when available. Results are reported as mean $\pm$ standard deviation over five seeds.

B. Same-receiver and receiver-shift evaluation

Same-receiver experiments use cycle-disjoint subsets of the same receiver. Receiver transfer is evaluated as R2 $\rightarrow$ R1 and R1 $\rightarrow$ R2, training on the source and testing on the target receiver.

Descriptor regression is restricted to paired waveforms with synchronized TX references and evaluated by mean absolute error (MAE).

C. Unseen-waveform distribution shift

Training and validation use the paired waveform identities

$$\mathcal{M}_{\text{seen}} = \{\text{PRBS, OFDM, OSDM, MFSK}\}. \quad (14)$$

Testing uses the non-paired identities

$$\mathcal{M}_{\text{unseen}} = \{\text{JANUS, BCSK, FH-BCSK, S2C-QPSK}\}. \quad (15)$$

The two sets are disjoint:

$$\mathcal{M}_{\text{seen}} \cap \mathcal{M}_{\text{unseen}} = \emptyset. \quad (16)$$

Because the raw waveform identities differ between training and testing, all signals are mapped to the shared label space

$$\mathcal{Y}_{\text{coarse}} = \{\text{tone/multicarrier, phase/spread-coded}\}.$$

The mapping is

$$\begin{aligned} \left\{ \begin{array}{l} \text{OFDM, OSDM,} \\ \text{MFSK, JANUS} \end{array} \right\} &\mapsto \text{tone/multicarrier,} \\ \left\{ \begin{array}{l} \text{PRBS, BCSK,} \\ \text{FH-BCSK, S2C-QPSK} \end{array} \right\} &\mapsto \text{phase/spread-coded.} \end{aligned} \quad (17)$$

This protocol evaluates generalization to unseen waveform identities according to spectral organization, rather than out-of-distribution detection or unknown-class rejection.

D. Compared representations

Three representations are evaluated. The **descriptor-space teacher reference** is a non-deployable MLP trained on privileged physical descriptors. The **RX-only CNN baseline** uses the same student architecture but is trained without privileged information. The **physics-distilled RX encoder** is trained using the cosine loss in Eq. (10). All frozen embeddings are evaluated with the same downstream linear models.

The teacher is treated as a training reference rather than an upper bound, since linear downstream models may exploit teacher and student embeddings differently.

E. Optimization and grid-search selection

Architecture and hyperparameters were selected using only cycle-disjoint training and validation partitions. The grid search evaluated

$$K \in \{32, 64, 128, 256\}, \quad (18)$$

cosine, mean-squared-error (MSE), and hybrid distillation losses, together with the following learning-rate/batch-size pairs:

$$\mathcal{H} = \{(5 \times 10^{-4}, 32), (10^{-3}, 16), (10^{-3}, 32), (10^{-3}, 64), (2 \times 10^{-3}, 32)\}. \quad (19)$$

Configurations were ranked by validation coarse-family accuracy, with mean descriptor MAE as a tie-breaker over five seeds. Test data were excluded.

The selected model uses a four-block CNN, $K = 128$, cosine alignment, Adam (10^{-3}), batch size 32, weight decay 10^{-4} , gradient clipping at 5.0, and early stopping with patience 15. Training statistics use only training data, and the configuration is fixed across bands and receivers.

Section VI-C evaluates descriptor removal, latent dimension, and distillation loss.

VI. RESULTS AND DISCUSSION

The experiments evaluate the consistency and transfer of frozen RX-only embeddings. Descriptor regression provides a paired-signal diagnostic, while unseen-waveform coarse-family classification is the main task.

A. Paired-signal descriptor diagnostic

Table I reports MAE under same-receiver and receiver-shift conditions. The physics-distilled RX encoder generally outperforms the baseline, approaches the teacher for `ds_rms`, and achieves the lowest `tail_50ms` error.

The distilled RX embedding achieves lower `cfo` and `tail_50ms` MAE than the teacher, which serves as a training reference rather than an upper bound.

B. Non-cooperative unseen-waveform evaluation

Because unseen waveforms lack synchronized TX references, transfer is evaluated with a linear probe in the shared coarse signaling-family space. Table II reports accuracy for seen and unseen identities.

TABLE II

COARSE SIGNALING-FAMILY ACCURACY USING FROZEN RX EMBEDDINGS. VALUES ARE MEAN $\pm$ STANDARD DEVIATION OVER FIVE SEEDS.

Representation	Seen waveform identities (%)		
	LF	MF	HF
RX-only CNN baseline	88.1 $\pm$ 1.2	90.4 $\pm$ 1.0	92.0 $\pm$ 0.9
Physics-distilled RX encoder	90.2 $\pm$ 1.0	92.1 $\pm$ 0.9	93.4 $\pm$ 0.7
Representation	Unseen waveform identities (%)		
	LF	MF	HF
RX-only CNN baseline	52.3 $\pm$ 2.6	59.8 $\pm$ 2.3	64.9 $\pm$ 2.1
Physics-distilled RX encoder	62.7 $\pm$ 2.4	68.9 $\pm$ 2.1	73.1 $\pm$ 1.8

For seen identities, distillation improves accuracy by 1.4–2.1 percentage points (pp), increasing to 8.2–10.4 pp for unseen identities, with slightly lower variability. This suggests less reliance on waveform-specific cues, although the evaluation is limited to a predefined two-class label space.

C. Component Ablation Analysis

Using the same cycle-disjoint splits and seeds, ablations evaluate descriptor removal, latent dimension, and distillation objective. Table III reports unseen-waveform accuracy after removing each descriptor.

For each band and seed, R1 and R2 accuracies were averaged before computing the mean and standard deviation across five seeds. Each entry therefore summarizes ten receiver-specific evaluations without treating paired receivers as independent.

TABLE III

EFFECT OF DESCRIPTOR REMOVAL ON UNSEEN-WAVEFORM ACCURACY (%), AVERAGED ACROSS RECEIVERS WITHIN EACH FREQUENCY BAND. VALUES ARE MEAN $\pm$ STANDARD DEVIATION.

Descriptor subset	Unseen-waveform accuracy (%)		
	LF	MF	HF
All five descriptors	62.7 $\pm$ 2.4	68.9 $\pm$ 2.1	73.1 $\pm$ 1.8
Remove peak	59.1 $\pm$ 2.6	65.2 $\pm$ 2.3	69.8 $\pm$ 2.0
Remove cfo	58.3 $\pm$ 2.5	64.1 $\pm$ 2.4	68.4 $\pm$ 2.1
Remove ds_rms	57.6 $\pm$ 2.7	63.3 $\pm$ 2.5	67.9 $\pm$ 2.2
Remove tail_50ms	56.8 $\pm$ 2.8	62.7 $\pm$ 2.6	67.1 $\pm$ 2.3
Remove env_kurtosis	60.2 $\pm$ 2.5	66.4 $\pm$ 2.2	70.9 $\pm$ 1.9

Removing `tail_50ms` causes the largest performance drop, reaching 5.9 pp in LF. Removing `ds_rms` or `cfo` also degrades performance, whereas `env_kurtosis` and `peak` have smaller effects.

Table IV examines the effect of embedding dimension and distillation loss on unseen-waveform accuracy.

TABLE I

MEAN ABSOLUTE ERROR FOR THREE PHYSICAL DESCRIPTORS UNDER SAME-RECEIVER AND RECEIVER-SHIFT EVALUATION. VALUES ARE MEAN $\pm$ STANDARD DEVIATION; STANDARD DEVIATIONS BELOW 0.01 ARE DISPLAYED AS 0.0 AFTER ROUNDING.

Descriptor	Representation	Same-receiver evaluation						Receiver-shift diagnostic					
		Test R1			Test R2			R2 $\rightarrow$ R1			R1 $\rightarrow$ R2		
		LF	MF	HF	LF	MF	HF	LF	MF	HF	LF	MF	HF
cfo (Hz)	Descriptor-space teacher reference	48.90 $\pm$ 1.6	47.80 $\pm$ 1.5	48.40 $\pm$ 1.6	50.60 $\pm$ 1.7	51.30 $\pm$ 1.8	51.00 $\pm$ 1.7	52.80 $\pm$ 1.9	53.60 $\pm$ 2.0	54.40 $\pm$ 2.0	54.90 $\pm$ 2.1	55.60 $\pm$ 2.1	55.20 $\pm$ 2.1
	RX-only CNN baseline	51.10 $\pm$ 1.8	50.20 $\pm$ 1.7	50.80 $\pm$ 1.8	53.00 $\pm$ 1.9	53.80 $\pm$ 2.0	53.40 $\pm$ 1.9	54.60 $\pm$ 2.0	55.30 $\pm$ 2.1	56.10 $\pm$ 2.2	56.70 $\pm$ 2.2	57.40 $\pm$ 2.3	57.10 $\pm$ 2.2
	Physics-distilled RX encoder	46.00 $\pm$ 1.4	45.10 $\pm$ 1.3	45.70 $\pm$ 1.4	47.70 $\pm$ 1.5	48.40 $\pm$ 1.6	48.00 $\pm$ 1.5	48.90 $\pm$ 1.7	49.40 $\pm$ 1.8	50.10 $\pm$ 1.9	50.60 $\pm$ 1.9	51.20 $\pm$ 2.0	50.90 $\pm$ 1.9
ds_rms (ms)	Descriptor-space teacher reference	0.14 $\pm$ 0.0	0.14 $\pm$ 0.0	0.14 $\pm$ 0.0	0.15 $\pm$ 0.0	0.16 $\pm$ 0.0	0.16 $\pm$ 0.0	0.16 $\pm$ 0.0	0.16 $\pm$ 0.0	0.17 $\pm$ 0.0	0.17 $\pm$ 0.0	0.17 $\pm$ 0.0	0.17 $\pm$ 0.0
	RX-only CNN baseline	0.18 $\pm$ 0.1	0.17 $\pm$ 0.1	0.17 $\pm$ 0.1	0.19 $\pm$ 0.1	0.20 $\pm$ 0.1	0.19 $\pm$ 0.1	0.19 $\pm$ 0.1	0.19 $\pm$ 0.1	0.20 $\pm$ 0.1	0.20 $\pm$ 0.1	0.21 $\pm$ 0.1	0.20 $\pm$ 0.1
	Physics-distilled RX encoder	0.15 $\pm$ 0.1	0.14 $\pm$ 0.1	0.14 $\pm$ 0.1	0.16 $\pm$ 0.1	0.16 $\pm$ 0.1	0.16 $\pm$ 0.1	0.16 $\pm$ 0.1	0.16 $\pm$ 0.1	0.17 $\pm$ 0.1	0.17 $\pm$ 0.1	0.17 $\pm$ 0.1	0.17 $\pm$ 0.1
tail_50ms (normalized)	Descriptor-space teacher reference	0.03 $\pm$ 0.0	0.03 $\pm$ 0.0	0.03 $\pm$ 0.0	0.03 $\pm$ 0.0	0.04 $\pm$ 0.0	0.04 $\pm$ 0.0	0.03 $\pm$ 0.0	0.03 $\pm$ 0.0	0.03 $\pm$ 0.0	0.04 $\pm$ 0.0	0.04 $\pm$ 0.0	0.04 $\pm$ 0.0
	RX-only CNN baseline	0.03 $\pm$ 0.0	0.03 $\pm$ 0.0	0.03 $\pm$ 0.0	0.03 $\pm$ 0.0	0.04 $\pm$ 0.0	0.04 $\pm$ 0.0	0.03 $\pm$ 0.0	0.03 $\pm$ 0.0	0.03 $\pm$ 0.0	0.04 $\pm$ 0.0	0.04 $\pm$ 0.0	0.04 $\pm$ 0.0
	Physics-distilled RX encoder	0.01 $\pm$ 0.0	0.01 $\pm$ 0.0	0.01 $\pm$ 0.0	0.01 $\pm$ 0.0	0.01 $\pm$ 0.0	0.01 $\pm$ 0.0	0.01 $\pm$ 0.0	0.01 $\pm$ 0.0	0.01 $\pm$ 0.0	0.01 $\pm$ 0.0	0.01 $\pm$ 0.0	0.01 $\pm$ 0.0

TABLE IV

EFFECT OF LATENT DIMENSION K AND DISTILLATION LOSS ON UNSEEN-WAVEFORM ACCURACY (%), AVERAGED ACROSS RECEIVERS WITHIN EACH FREQUENCY BAND.

Configuration	Unseen-waveform accuracy (%)		
	LF	MF	HF
$K = 32$, cosine	57.8 $\pm$ 2.5	64.1 $\pm$ 2.3	68.3 $\pm$ 2.1
$K = 64$, cosine	60.4 $\pm$ 2.4	66.7 $\pm$ 2.2	71.2 $\pm$ 1.9
$K = 128$, cosine	62.7 $\pm$ 2.4	68.9 $\pm$ 2.1	73.1 $\pm$ 1.8
$K = 256$, cosine	62.1 $\pm$ 2.5	68.2 $\pm$ 2.2	72.4 $\pm$ 1.9
$K = 128$, MSE	58.9 $\pm$ 2.6	65.3 $\pm$ 2.4	69.7 $\pm$ 2.1
$K = 128$, hybrid	60.8 $\pm$ 2.5	67.1 $\pm$ 2.2	71.5 $\pm$ 2.0

Cosine distillation performs best across all bands, with the highest accuracy at $K = 128$. The descriptor-removal results further highlight the contribution of multipath-related information.

D. Scope and limitations

The selected configuration is supported only by the evaluated search space and SODAC conditions. Descriptor interactions were not exhaustively examined, and the task addresses unseen identities within a shared label space rather than unknown-waveform detection.

VII. CONCLUSION

This paper presented a physics-guided RX-only framework for UWA communications using LUPI and knowledge distillation. A teacher learns from paired TX–RX descriptors, while a CNN student uses only RX spectrograms at inference. Distillation improves descriptor diagnostics and unseen-waveform transfer by 8.2–10.4 pp. Ablations identify `tail_50ms`, `ds_rms`, and `cfo` as the most relevant descriptors, with cosine distillation and $K = 128$ performing best. Future work will address additional environments, uncertainty estimation, and time-varying channels.

REFERENCES

- [1] M. Stojanovic, "On the Relationship Between Capacity and Distance in an Underwater Acoustic Communication Channel," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 11, no. 4, pp. 34–43, 2007, doi: 10.1145/1347364.1347373.
- [2] P. Qarabaqi and M. Stojanovic, "Statistical Characterization and Computationally Efficient Modeling of a Class of Underwater Acoustic Communication Channels," *IEEE Journal of Oceanic Engineering*, vol. 38, no. 4, pp. 701–717, 2013, doi: 10.1109/OJOE.2013.2278787.
- [3] N. Morozs, W. Gorma, B. T. Henson, L. Shen, P. D. Mitchell, and Y. V. Zakharov, "Channel Modeling for Underwater Acoustic Network Simulation," *IEEE Access*, vol. 8, pp. 136151–136175, 2020, doi: 10.1109/ACCESS.2020.3011620.
- [4] M. B. Porter, *The BELLHOP Manual and User's Guide*. Heat, Light, and Sound Research, Inc., 2011.
- [5] P. van Walree, R. Otnes, B. Tomasi, B. Henriksen, A. Våhlin, J.-B. Danre, Ø. Bergh, F.-X. Socheleau, and M. Chitre, "Descriptor: The SFI Smart Ocean Dataset for Acoustic Communications (SODAC)," *IEEE Data Descriptions*, vol. 2, pp. 218–223, 2025, doi: 10.1109/IEEE-DATE.2025.3577998.
- [6] S. Hershey, S. Chaudhuri, D. P. W. Ellis, J. F. Gemmeke, A. Jansen, R. C. Moore, M. Plakal, D. Platt, R. A. Saurous, B. Seybold, and M. Slaney, "CNN Architectures for Large-Scale Audio Classification," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)*, 2017, doi: 10.1109/ICASSP.2017.7952132.
- [7] C. Shen, Y. Li, X. Zhang, and Y. Yang, "Deep Learning-Based Underwater Acoustic Signal Classification," *Applied Acoustics*, vol. 167, p. 107383, 2020, doi: 10.1016/j.apacoust.2020.107383.
- [8] Z. Li, W. Li, K. Sun, D. Fan, and W. Cui, "Recent Progress on Underwater Wireless Communication Methods and Applications," *Journal of Marine Science and Engineering*, vol. 13, no. 8, p. 1505, 2025, doi: 10.3390/jmse13081505.
- [9] V. Vapnik and A. Vashist, "Learning Using Privileged Information," *Neural Networks*, vol. 22, no. 5–6, pp. 544–557, 2009, doi: 10.1016/j.neunet.2009.06.042.
- [10] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," *arXiv preprint arXiv:1503.02531*, 2015.
- [11] D. Lopez-Paz, L. Bottou, B. Schölkopf, and V. Vapnik, "Unifying Distillation and Privileged Information," in *International Conference on Learning Representations (ICLR)*, 2016.
- [12] F. Tung and G. Mori, "Similarity-Preserving Knowledge Distillation," in *Proc. IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.
- [13] I. A. Tasadduq and M. Rashid, "Toward Intelligent Underwater Acoustic Systems: Systematic Insights into Channel Estimation and Modulation Methods," *Electronics*, vol. 14, no. 15, p. 2953, 2025, doi: 10.3390/electronics14152953.
- [14] E. Panteli, P. E. Santos, and N. Humphrey, "AquaSignal: An Integrated Framework for Robust Underwater Acoustic Analysis," *arXiv preprint arXiv:2505.14285*, 2025, doi: 10.48550/arXiv.2505.14285.
- [15] F. Ardizzone, F. Giuriso, and S. Tomasin, "Advantage-Distillation Strategies with Side Information for Underwater Acoustic Channels," in *Proc. 10th Convention of the European Acoustics Association, Forum Acusticum 2023*, Turin, Italy, 2023.
- [16] S. Yang, G. B. Deane, J. C. Preisig, N. C. Seviktekin, J. W. Choi, and A. C. Singer, "On the Reusability of Postexperimental Field Data for Underwater Acoustic Communications R&D," *IEEE Journal of Oceanic Engineering*, vol. 44, no. 4, pp. 912–931, 2019, doi: 10.1109/OJOE.2019.2925921.